ROUND-OFF ERRORS IN MEDIAEVAL TABLES

by

M.A. Stephens

TECHNICAL REPORT No. 466

MARCH 16, 1993

Prepared Under Contract
N00014-92-J-1264 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

ROUND-OFF ERRORS IN MEDIAEVAL TABLES

by

M.A. Stephens

TECHNICAL REPORT No. 466

MARCH 16, 1993

Prepared Under Contract

N00014-92-J-1264 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

Professor Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305-4065

ROUND-OFF ERRORS

IN

MEDIAEVAL

TABLES

by

M.A. Stephens

Summary

The distribution of error is considered when a function y of x is rounded, and when x is uniformly distributed. The example discussed is $y = \sin x$, and it is thought that the round-off error might be nearly uniformly distributed.

The non-uniformity is very small, and the sample size needed to detect this by the A^2 statistic is examined. The study is of interest in the examination of ancient and mediaeval tables.

Key Words: distribution of error: goodness-of-fit; tables of functions;

1 INTRODUCTION

This problem was brought to the Statistical Consulting Service at Simon Fraser University by G. van Brummelen, a graduate student in the History of Mathematics. It contains interesting probabilistic features and a statistical application which appear to be worth recording.

The problem concerns the distribution of error ϵ when a value of y, a function of x, is rounded, say to 2 d.p. The distribution is the sum of many terms, and at first sight it may appear to be approximately uniform: this brings in the statistical application, to examine how one would detect that it is non-uniform.

Mr. van Brummelen describes the origin of the problem as follows:

> Many ancient and medieval astronomical treatises contain numerical tables which allow the reader to calculate planetary positions and related phenomenae. The formulae implicit in these tables are given, but the errors in the tabular values do not usually reflect what one would expect from a straightforward computation. This may be due to the use of interpolation or other timesaving techniques, or to varying levels of rounding. In order to determine the calculation methods used by the author of a table, I am developing (have developed) several numerical and statistical tests. These are designed to search for interpolation grids, check for dependence of one table on another, and find an error distribution given an hypothesized calculation method, for example. Many of the tests require the assumption that the error caused by rounding a set of computed values to some level is nearly uniformly distributed. It is this assumption that I wish to verify.

2. DISTRIBUTION OF ERROR for $y = \sin x$.

We examine the distribution of round-off error for the function $y = \sin x$; the value of x is considered to be uniformly distributed in the interval $0 \le x \le \pi/2$. Suppose y is rounded to accuracy Δ; in what follows we assume $\Delta = 0.01$. Then the error in y is found as follows. Suppose y is rounded to $i\Delta$, for $i=1,\ldots,n-1$; the true value must have been y^*, such that $y_1 < y^* \le y_2$ where $y_1 = (i\Delta - \Delta/2)$ and $y_2 = (i\Delta + \Delta/2)$. These correspond to $x_{1i} = \sin^{-1} y_1$ and $x_{2i} = \sin^{-1} y_2$; call the interval $(x_{1i} < x \le x_{2i})$ the i-th interval. The error in y is

$$\epsilon = y - \sin x \tag{1}$$

Suppose $F(t)$ is the distribution of ϵ: that is, $F(t) = P(\epsilon < t)$. The contribution to $F(t)$ from the i-th interval, for $1 < i < n-1$, is

$$F_i(t) = \frac{2}{\pi} (\sin^{-1}(i\Delta + \Delta/2) - \sin^{-1}(i\Delta - t)) , \quad -\frac{\Delta}{2} \le t \le \frac{\Delta}{2} . \tag{2}$$

The top and bottom intervals are special cases. For $0 \le x \le \sin^{-1}(\Delta/2)$ the error is negative, and the contribution to $F(t)$ is $F_0(t) = 2(\sin^{-1}(\Delta/2) - \sin^{-1}(t))/\pi$; for $\sin^{-1}(1-\Delta/2) \le x \le \pi/2$, the error is positive and the contribution to $F(t)$ is $F_n(t) = 2(\pi/2 - \sin^{-1}(1-t))/\pi$. When these are put together we have finally, with $n\Delta = \pi/2$,

$$F(t) = \frac{2}{\pi} \left[\sum_{i=1}^{n-1} (\sin^{-1}(i\Delta+\Delta/2) - \sin^{-1}(i\Delta-t)) + \sin^{-1} \Delta/2 + \sin^{-1}(t) \right] ,$$

$$- \Delta/2 \le t \le 0;$$

and

$$F(t) = \frac{2}{\pi} \left[\sum_{i=1}^{n-1} (\sin^{-1}(i\Delta+\Delta/2) - \sin^{-1}(i\Delta-t)) + \frac{\pi}{2} - \sin^{-1}(1-t) + \sin^{-1} \Delta/2 \right] ,$$

$$0 \le t \le \Delta/2 .$$

The density is

$$f(t) = \frac{2}{\pi}\left[\sum_{i=1}^{n-1} 1/(1 - (i\Delta-t)^2)^{1/2} + 1/(1-t^2)^{1/2}\right], \quad -\Delta/2 \le t \le 0;$$

and

$$f(t) = \frac{2}{\pi}\left[\sum_{i=1}^{n-1} 1/(1-(i\Delta-t)^2)^{1/2} + 1/(2t-t^2)^{1/2}\right], \qquad 0 < t \le \Delta/2;$$

The density has an infinite value as t approaches zero from above. Thus, it is certainly not uniform, but through much of the range it will be close to uniform. Table 1 gives values of $F(\epsilon)$ and $f(\epsilon)$ for a range of values of ϵ, when $\Delta = 0.01$. Figures 1a, 1b are plots of $F(\epsilon)$ and $f(\epsilon)$ from Table 1, and Figures 2a, 2b are similar plots on a larger scale, to show the sharp change in density at $\epsilon=0$.

3. THE DETECTION OF NON-UNIFORMITY

The following statistical problem can then be posed. Suppose $U(a,b)$ denotes the uniform distribution between a, b, and suppose a sample of size N is taken from $F(t)$. How large must N be in order to reject

$$H_0\text{: the errors } \epsilon \text{ are } U(-\Delta/2, \Delta/2)\ ?$$

The size of N will clearly depend on the statistic used: we have examined a statistic which is generally accepted to be powerful for such a test, namely the EDF statistic A^2 (for the definition and tables, see Stephens, 1986).

Table 2 gives the number of 100 Monte Carlo samples which were detected as significant by this statistic, using samples of size N. The percentages are given for several test sizes α. Three sets of samples of size 2000, and two of sizes 5000 and 10000 were included to show the variability in power of A^2 to detect the non-uniformity. The table shows that even with 2000 values, a 5% test would detect this delicate departure from uniformity only about 20 times in 100; the sample size must go to 10,000 to find a power of over 85%. Thus one can suppose the error distribution will appear uniform to many observers and can probably be treated as such for many statistical purposes.

REFERENCES

Stephens, M.A. (1986) Tests based on EDF statistics. Chapter 4 in "Goodness-of-Fit Techniques" (R.B. d'Agostino and M.A. Stephens, eds.) Marcell Dekker: New York.

Table 1

Values of distribution and density of ε

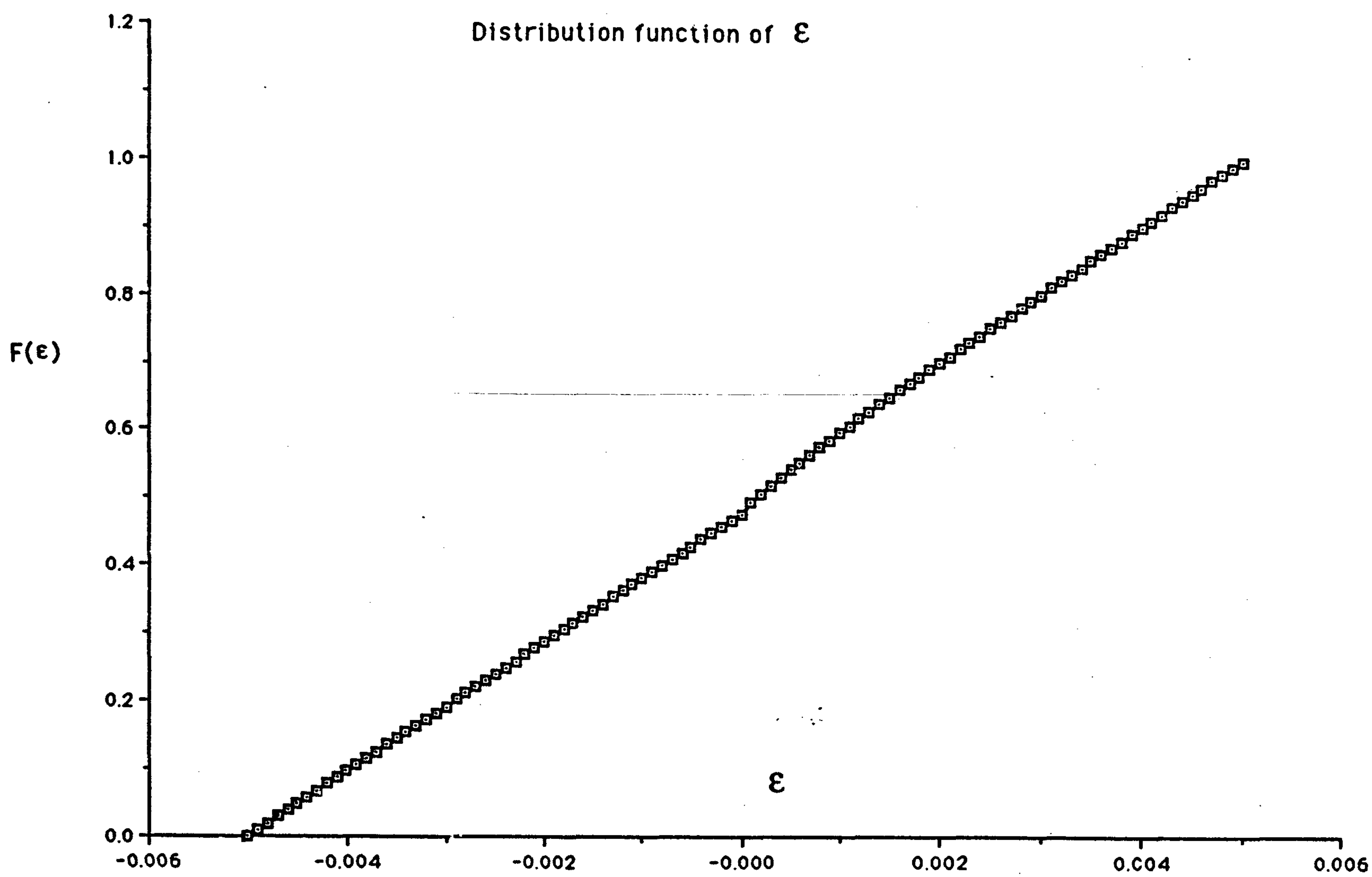
e	F(e)	f(e)
-0.0050	0.0000	97.28
-0.0047	0.0243	97.03
-0.0045	0.0485	96.80
-0.0042	0.0727	96.57
-0.0040	0.0968	96.36
-0.0038	0.1209	96.15
-0.0035	0.1449	95.95
-0.0033	0.1688	95.76
-0.0030	0.1928	95.58
-0.0028	0.2166	95.40
-0.0025	0.2405	95.23
-0.0023	0.2642	95.06
-0.0020	0.2880	94.90
-0.0018	0.3117	94.74
-0.0015	0.3354	94.59
-0.0013	0.3590	94.44
-0.0010	0.3826	94.29
-0.0008	0.4061	94.15
-0.0005	0.4297	94.01
-0.0003	0.4531	93.87
0.0000	0.4766	93.74
0.0002	0.5141	121.45
0.0005	0.5432	112.98
0.0007	0.5709	109.16
0.0010	0.5979	106.84
0.0012	0.6244	105.22
0.0015	0.6506	103.99
0.0017	0.6764	103.01
0.0020	0.7021	102.21
0.0022	0.7275	101.52
0.0025	0.7528	100.92
0.0027	0.7780	100.40
0.0030	0.8030	99.92
0.0032	0.8280	99.50
0.0035	0.8528	99.11
0.0037	0.8775	98.75
0.0040	0.9022	98.41
0.0042	0.9267	98.10
0.0045	0.9512	97.81
0.0047	0.9756	97.54
0.0050	1.0000	97.28

TABLE 2

The tablcs gives the number of Monte Carlo samples, each of size N, which give a significant value of A^2 at level α. The number of Monte Carlo samples generated, for each N, was 100.

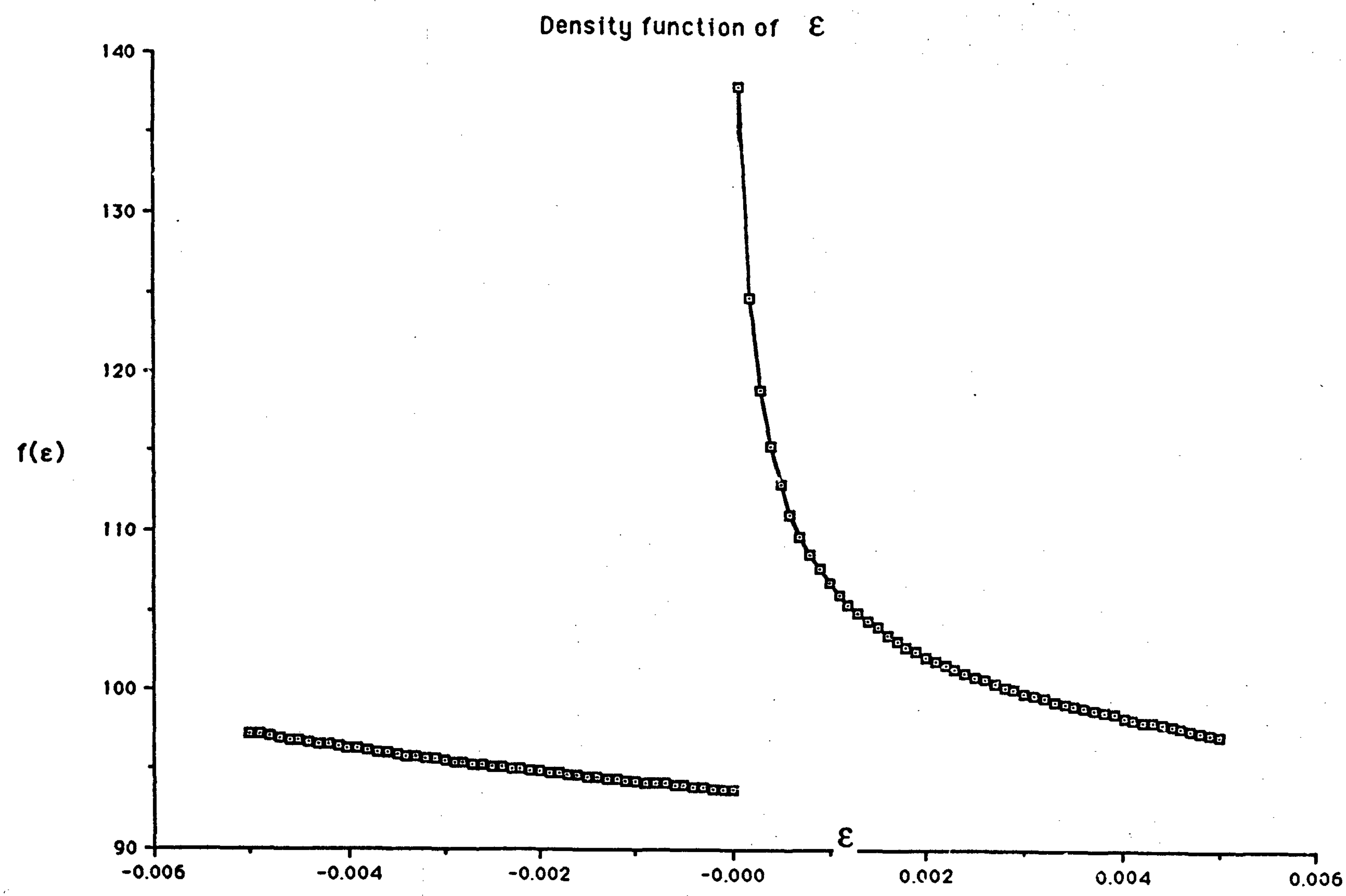
N \ α:	0.25	0.10	0.05	0.01
2000	60	31	16	4
2000	56	32	23	6
2000	59	36	21	6
5000	92	77	67	25
5000	89	67	50	23
10000	100	96	88	69
10000	100	99	91	63

Fig. 1a.



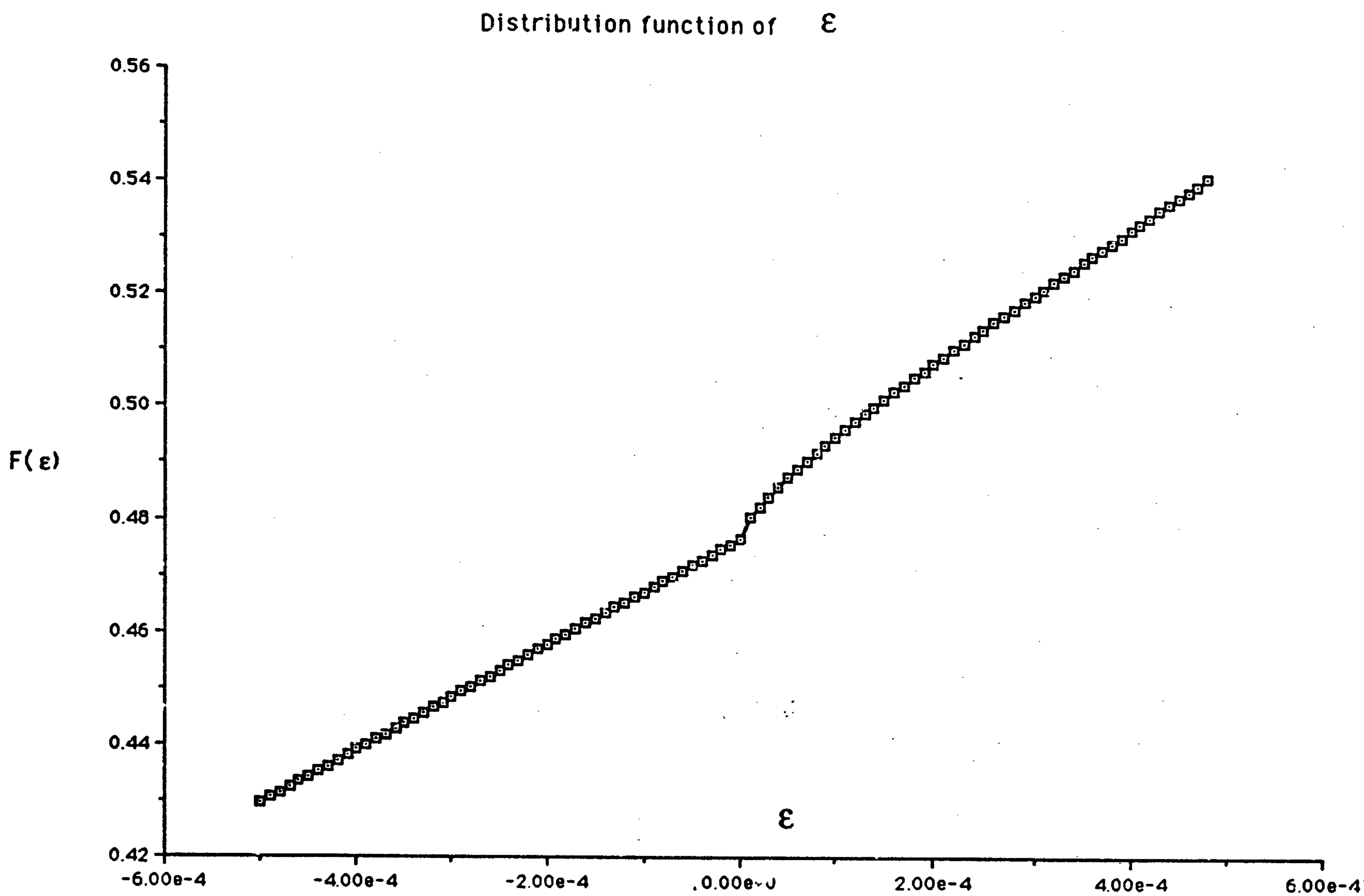
Distribution function of ε
F(ε)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.006
-0.004
-0.002
-0.000
0.002
0.004
0.006
ε

Fig. 1b.



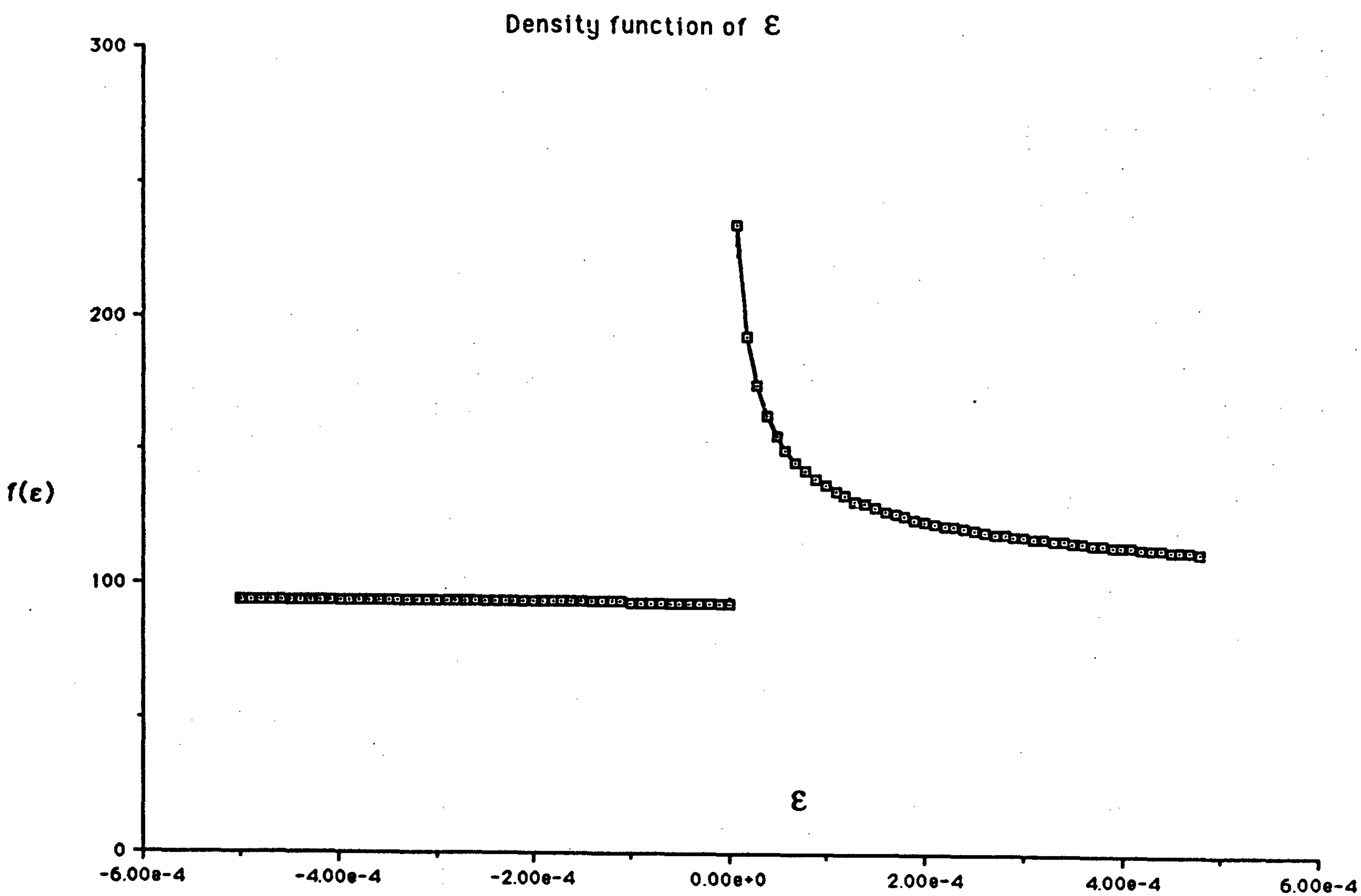
Density function of ε
f(ε)
140
130
120
110
100
90
-0.006
-0.004
-0.002
-0.000
0.002
0.004
0.006
ε

Fig. 2a.



Distribution function of ε
F(ε)
0.56
0.54
0.52
0.50
0.48
0.46
0.44
0.42
ε
-6.00e-4
-4.00e-4
-2.00e-4
0.00e-0
2.00e-4
4.00e-4
6.00e-4

Fig. 2b.



Density function of ε
300
200
100
0
f(ε)
ε
-6.00e-4
-4.00e-4
-2.00e-4
0.00e+0
2.00e-4
4.00e-4
6.00e-4

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 466	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Round-off Errors in Mediaeval Tables		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)		8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 111		12. REPORT DATE March 16, 1993
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

distribution of error: goodness-of-fit; tables of functions;

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

Summary

The distribution of error is considered when a function y of x is rounded, and when x is uniformly distributed. The example discussed is $y = \sin x$, and it is thought that the round-off error might be nearly uniformly distributed.

The non-uniformity is very small, and the sample size needed to detect this by the A^2 statistic is examined. The study is of interest in the examination of ancient and mediaeval tables.

END

FILMED

DATE: 4-93

DTIC